

# Noise properties of CT images reconstructed by use of constrained total-variation, data-discrepancy minimization

**Rose, Sean; Andersen, Martin S.; Sidky, Emil Y.; Pan, Xiaochuan**

*Published in:*
Medical Physics

*Link to article, DOI:*
10.1118/1.4914148

*Publication date:*
2015

*Document Version*
Publisher's PDF, also known as Version of record

Medical Physics
The International Journal of Medical Physics Research and Practice

# Noise properties of CT images reconstructed by use of constrained total-variation, data-discrepancy minimization

Sean Rose, Martin S. Andersen, Emil Y. Sidky, and Xiaochuan Pan

Citation: Medical Physics **42**, 2690 (2015); doi: 10.1118/1.4914148
View online: http://dx.doi.org/10.1118/1.4914148
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/42/5?ver=pdfcov
Published by the American Association of Physicists in Medicine

# Noise properties of CT images reconstructed by use of constrained total-variation, data-discrepancy minimization

Sean Rose[a)]
*Department of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637*

Martin S. Andersen[b)]
*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby 2800, Denmark*

Emil Y. Sidky[c)] and Xiaochuan Pan[d)]
*Department of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637*

(Received 1 September 2014; revised 8 December 2014; accepted for publication 8 January 2015; published 4 May 2015)

**Purpose:** The authors develop and investigate iterative image reconstruction algorithms based on data-discrepancy minimization with a total-variation (TV) constraint. The various algorithms are derived with different data-discrepancy measures reflecting the maximum likelihood (ML) principle. Simulations demonstrate the iterative algorithms and the resulting image statistical properties for low-dose CT data acquired with sparse projection view angle sampling. Of particular interest is to quantify improvement of image statistical properties by use of the ML data fidelity term.
**Methods:** An incremental algorithm framework is developed for this purpose. The instances of the incremental algorithms are derived for solving optimization problems including a data fidelity objective function combined with a constraint on the image TV. For the data fidelity term the authors, compare application of the maximum likelihood principle, in the form of weighted least-squares (WLSQ) and Poisson-likelihood (PL), with the use of unweighted least-squares (LSQ).
**Results:** The incremental algorithms are applied to projection data generated by a simulation modeling the breast computed tomography (bCT) imaging application. The only source of data inconsistency in the bCT projections is due to noise, and a Poisson distribution is assumed for the transmitted x-ray photon intensity. In the simulations involving the incremental algorithms an ensemble of images, reconstructed from 1000 noise realizations of the x-ray transmission data, is used to estimate the image statistical properties. The WLSQ and PL incremental algorithms are seen to reduce image variance as compared to that of LSQ without sacrificing image bias. The difference is also seen at few iterations—short of numerical convergence of the corresponding optimization problems.
**Conclusions:** The proposed incremental algorithms prove effective and efficient for iterative image reconstruction in low-dose CT applications particularly with sparse-view projection data.  [http://dx.doi.org/10.1118/1.4914148]

Key words: computed tomography, iterative image reconstruction, total variation, noise properties, incremental algorithms

## 1. INTRODUCTION

In iterative image reconstruction (IIR), there can be a large disconnect between practical iterative algorithms and the optimization problems that motivate their design. Particularly for image reconstruction from sparse-view CT data with its associated ill-conditioned linear system model, the number of required iterations for accurate solvers of relevant optimization problems can be much greater than 1000. When iteration numbers are this large, the trajectory of the image estimates can be quite important because practical application of IIR dictates iteration numbers on the order of ten or one hundred—well short of convergence.

A common strategy for obtaining useful images rapidly is to employ algorithms that process the data sequentially.[1,2] In particular, for sparse-view CT we have been developing the adaptive-steepest-descent–projection-onto-convex-sets (ASD–POCS) algorithm,[3] which is sequential in that it employs ART for the data agreement step. This algorithm has been shown to yield useful images at low iteration numbers.[4] The version of ASD-POCS reported in Ref. 3 is designed to solve the data-discrepancy-constrained total-variation (TV) minimization problem. For low-dose CT, where the data noise is relatively high, we have adapted ASD-POCS for accurate solution of TV constrained, data-discrepancy minimization.[5] While this algorithm was effective for its purpose, it required at least 100 iterations and the algorithm had five control parameters, *four* of which required tuning. With the present incremental framework useful images can be obtained from the TV-constrained optimization problem with as few as twenty iterations, and only *one* algorithm parameter needs to be tuned.

Recently, an incremental framework[6,7] has been developed from which sequential iterative algorithms can be derived that both yield useful images at low iteration numbers and converge to the solution of a designed optimization problem. The reason why such a framework can be helpful for IIR algorithm development is that many design principles such as maximum entropy, maximum likelihood (ML), and sparsity exploitation are a form of optimization. It is not clear that truncating the iteration of the optimization problem solver will yield images that reflect the intentions of the designed optimization problem. With the incremental framework, where initial convergence is rapid, there may be a stronger link between early image estimates and the solution to the designed optimization.

In this work, we investigate the use of TV-constrained data-discrepancy minimization for sparse-view image reconstruction from noisy CT data. The use of the TV seminorm is motivated by selecting images with a sparse gradient magnitude image (GMI), and TV is known to be effective in reducing artefacts due to view angle under-sampling. Introducing this seminorm in the form of a constraint allows us to compare different data fidelity objective functions on an equal footing. In particular, we investigate the use of data fidelity terms derived from the maximum likelihood principle. Simulated CT data are generated, modeling the low-dose conditions of breast CT, using a Poisson statistics noise model for the transmitted intensity. Image reconstruction is performed with incremental algorithms, which constrain the image TV and minimize: (1) unweighted least-squares (LSQ), (2) weighted least-squares (WLSQ), and (3) Poisson-likelihood (PL) motivated objective functions. The WLSQ objective function is designed to approximate that of PL. The comparisons are performed on estimating statistical properties of the reconstructed images from 1000 noise realizations of the simulated transmission intensity data. In Sec. 2, the theoretical background for the incremental algorithms is presented along with the algorithms themselves. In Sec. 3, extensive simulations are performed comparing the incremental algorithms for IIR from low-dose CT data. In Sec. 4, we discuss practical aspects on the use of the various incremental algorithms focusing also on the application of the maximum-likelihood principle.

## 2. DATA MODEL AND INCREMENTAL ALGORITHMS

### 2.A. Data model

We employ a generic linear model for x-ray projection

$$Ax = b, \tag{1}$$

where $x \in \Re^n$ represents an image; $b \in \Re^m$ is a sinogram; and $A \in \Re^{m\times n}$ denotes x-ray projection. For the present study, we consider a CT data model where the transmitted x-ray intensity follows a Poisson distribution. Let $N_0$ be the number of photons incident on the object for each ray and consider modeling the number of photons incident on the detector in the $i$th ray, $\mathbf{y}_i$, as a Poisson random variable with mean $N_0 p_i(x)$. We then have

$$\Pr(\mathbf{y}_i = y_i) = \frac{(N_0 p_i(x))^{y_i}}{y_i!} e^{-N_0 p_i(x)}. \tag{2}$$

For transmission tomography, we take

$$p_i(x) = \exp(-a_i^T x), \tag{3}$$

where $a_i$ is the $i$th row of the projection matrix $A$. We note that it is possible that the measured transmitted intensity $\mathbf{y}_i$ for a particular ray could be zero, and in this case, the measurement does not contain reliable information on the object and it is simply removed from the data set and $A$ is modified by removing the corresponding row $a_i$. For the scanning configurations considered here, the projection view angle is under-sampled; namely, there are too few views for Eq. (1) to specify a unique solution. In designing an IIR algorithm for this noise model and configuration, we consider data fidelities motivated by maximum likelihood combined with a constraint capping the image TV.

### 2.B. Optimization problems

#### 2.B.1. TV-constrained PL maximization

We formulate a convex optimization problem based on the PL function. Given $m$ measurements $\mathbf{N_i}$, assuming that each random variable $\mathbf{N_i}$ is independent, the log likelihood function is given by

$$\mathcal{L}(x|y) = \ln\left(\prod_{i=1}^{m} \Pr(\mathbf{y}_i = y_i)\right), \tag{4}$$

$$\propto \sum_{i=1}^{m} \left(-y_i a_i^T x - N_0 \exp(-a_i^T x)\right). \tag{5}$$

As opposed to maximizing this expression over $x$, we minimize its negative subject to a TV constraint. Thus, given a data vector of x-ray transmission measurements $y \in \Re^m$, our goal is to solve the convex optimization problem,

$$\min_x \sum_{i=1}^{m} \left(y_i a_i^T x + N_0 \exp(-a_i^T x)\right)$$
$$\text{such that } \mathrm{TV}(x) \le \gamma. \tag{6}$$

The parameter $\gamma$ is the TV constraint parameter.

The TV function is given by

$$\mathrm{TV}(x) = \|Dx\|_{1,2} = \sum_{i=1}^{n} \|D_i x\|_2. \tag{7}$$

where $D_i \in \Re^{d\times n}$ is a finite differencing matrix approximating the gradient of the $d$-dimensional image at the $i$th pixel and $D \in \Re^{nd\times n}$ is given by

$$D = \begin{pmatrix} D_1 \\ \vdots \\ D_n \end{pmatrix}. \tag{8}$$

Including TV in Eq. (6) this way has a couple of advantages: comparisons using other data fidelity terms or different

sampling schemes can be made on an equal footing, and simulation studies are simplified in that the TV of the test phantom is available to use as $\gamma$.

### 2.B.2. TV-constrained WLSQ

When large numbers of photons are detected, the probability distribution in Eq. (2) can be approximated by a multivariate Gaussian with mean equal to the variance. Working with the negative logarithm of the mean transmission data,

$$b = -\ln\left(\frac{y}{N_0}\right),$$

the transformed data are approximately distributed

$$\Pr(\mathbf{b} = b) \propto \exp\left(-\frac{1}{2}(Ax-b)^T\Sigma^{-1}(Ax-b)\right), \tag{9}$$

where the covariance $\Sigma$ is diagonal with variance equal to one over the mean transmitted x-ray intensity,

$$\Sigma = \mathrm{diag}\left(\frac{1}{N_0\exp(-Ax)}\right). \tag{10}$$

To incorporate this likelihood into an optimization, we consider the general TV-constrained WLSQ problem,

$$\min_x \frac{1}{2}(Ax-b)^T W(Ax-b)$$
$$\text{such that } \mathrm{TV}(x) \le \gamma. \tag{11}$$

The matrix $W \in \Re^{m\times m}$ is a diagonal weighting matrix, and we refer to the individual diagonal elements as $w_i$ for $i = 1,\ldots, m$. The maximum likelihood data fidelity is achieved by setting the weights to $\Sigma^{-1}$,

$$W = \mathrm{diag}(N_0\exp(-Ax)).$$

We note that the TV-constrained LSQ problem corresponds to the case where the diagonal elements of $W$ are set to one.

## 2.C. Incremental algorithms

A class of algorithms known as incremental proximal gradient methods (IPG) have been proposed[6] to efficiently solve large-scale, convex optimization problems. The incremental framework of interest applies to an objective function $f : \Re^n \mapsto \Re$ written as the sum of $S$ terms in the following form:

$$f(x) = \sum_{i=1}^{S}(g_i(x) + h_i(x)), \tag{12}$$

where $g_i, h_i : \Re^n \mapsto \Re$ are convex. The optimization problems of Eqs. (6) and (11) fit this form in this sense: each of the data fidelity functions consist of sums of terms depending only on single measurements and the TV constraint, coded as an indicator function, is represented as one term. In these incremental methods, one takes first-order steps seeking to minimize each $g_i$ and $h_i$ individually. Empirically, fast initial convergence rates have been observed with such methods.[8]

More recently, relaxed versions of the IPG methods have been proposed.[7] We present the $k$th update steps for the one of the relaxed incremental proximal gradient (R-IPG) methods for minimization of a function $f$ as in Eq. (12) subject to a convex constraint set $X$ here,

$$\text{(R-IPG1)} \qquad \begin{aligned} q_k &= \mathrm{prox}_{t_k g_{i_k}}(x_k), \\ z_k &= q_k - t_k\widetilde{\nabla}h_{i_k}(q_k), \\ x_{k+1} &= \mathcal{P}_X(\rho z_k + (1-\rho)x_k), \end{aligned}$$

where $\mathcal{P}_X : \Re^n \mapsto \Re^n$ is the projection onto the convex set $X$ and the proximal mapping is given by

$$\mathrm{prox}_{t_k g_{i_k}}(x) = \arg\min_u \left\{ g_{i_k}(u) + \frac{\|x-u\|_2^2}{2t_k} \right\}. \tag{13}$$

Note the arg min operation returns a unique value since the function in its argument is strongly convex. The symbol $\widetilde{\nabla}h_i(w_k)$ represents a subgradient of $h_i$ at $w_k$. The value of the relaxation parameter $\rho$ must lie in the interval (0,2) to ensure convergence.

The value of $i_k$ can be chosen in a cyclic fashion such as $i_k = k\%S$—where we have used % to denote the modulus operation—or in a randomized manner. We refer the interested reader to Ref. 6 for more information about convergence rates in both cases. The step-size $t_k$ must be chosen such that

$$\sum_{k=1}^{\infty} t_k = \infty, \tag{14}$$
$$\lim_{k\to\infty} t_k = 0, \tag{15}$$

to ensure the algorithm does not end up in a limit cycle yet can still travel far enough to reach the optimal set.

Technically, the optimization problems (6) and (11) do not meet the criteria for convergence proof of R-IPG1 because the indicator function coding the TV constraint is not Lipschitz continuous. We make a modification to these optimization problems, which does not affect the implementation of R-IPG1 when the iteration is truncated. We write both optimizations in the form,

$$\min_x \{F(x) + \lambda d_C(x)\}, \tag{16}$$

where $F(x)$ represents either of the data fidelity objective functions, $\lambda$ is a parameter, and $d_C(x)$ is the distance from the point $x$ to the TV constraint set $C$,

$$C \equiv \{x : \mathrm{TV}(x) \le \gamma\}.$$

Formulating the optimization problems in this way, the objective functions are now Lipschitz continuous so that convergence of R-IPG1 is guaranteed. The role of $\lambda$ will be addressed shortly.

We derive an instance of R-IPG1 for solving TV-constrained PL maximization by defining

$$\begin{aligned} &g_i(x) = y_i a_i^T x + N_0\exp(-a_i^T x) \text{ for } i = 1,\ldots, m \\ &g_{m+1}(x) = \lambda d_C(x), \\ &h_i(x) = 0 \text{ for } i = 1,\ldots, m+1 \end{aligned} \tag{17}$$

and instances of TV-constrained WLSQ and LSQ (if the weights are set to one) by defining

$$g_i(x) = \frac{w_i}{2}(a_i^T x - b_i)^2 \text{ for } i = 1,\ldots, m,$$
$$g_{m+1}(x) = \lambda d_C(x),$$
$$h_i(x) = 0 \text{ for } i = 1,\ldots, m+1. \quad (18)$$

For implementation of these algorithms, we need to compute the various proximal mappings.

We start with the proximal mapping of the distance function $d_C(x)$ because this step determines the final form of R-IPG1 used in this work. It is shown in Appendix A that

$$\text{prox}_{t_k g_{m+1}}(x_k) = \begin{cases} x_k + t_k\lambda\dfrac{(\mathcal{P}_X(x_k) - x_k)}{d_C(x_k)} & \text{if } d_C(x_k) \geq \lambda t_k \\ \mathcal{P}_C(x_k) & \text{else} \end{cases},$$

where $\mathcal{P}_C(\cdot)$ denotes projection on the TV constraint set $C$. We note that the projection in the second case corresponds to the proximal mapping of the original indicator function for $C$. The modification due to changing this constraint to a distance function comes in the first case when $d_C(x_k) \geq \lambda t_k$. For truncated iteration and $t_k > 0$ and $\lambda$ can be chosen large enough that this inequality is not satisfied over the range of $k$. In practice, we always take this update step to be $\mathcal{P}_C(x_k)$ and there is no change in the instance of R-IPG1 in switching the TV constraint to a distance function. We can numerically calculate the projection $\mathcal{P}_C(x_k)$ using an efficient primal–dual algorithm such as that of Chambolle and Pock (CP).[9,10]

#### 2.C.1. Incremental algorithms for TV-constrained PL maximization, WLSQ, and LSQ

The desired form of R-IPG1 is derived by setting $\lambda$ large enough

$$\lambda > d_C(x_k)/t_k,$$

yielding the following R-IPG1 update steps

$$x_{k+1} = \begin{cases} \text{prox}_{t_k g_{i_k}}(x_k) & \text{if } k\%(m+1) \neq 0 \\ \rho\mathcal{P}_C(x_k) + (1-\rho)x_k & \text{else} \end{cases}. \quad (19)$$

The step in first case is derived from the proximal mapping of the data fidelity terms for PL maximization, WLSQ, and LSQ. For this update, the projection onto the convex set $X$ does not appear because we do not include additional constraints. Adding simple constraints, such as bound constraints, can be easily incorporated by specifying the appropriate constraint set $X$ for the image $x$. In practice, such constraints need to be simple because they are performed at every update step $k$; for the data fidelity steps the index $k$ specifies each of the terms in the corresponding objective and complex projections after each data term will prohibitively slow down the R-IPG1 algorithm. This is why the projection onto $C$ is taken as an incremental step for $k\%(m+1) = 0$ instead of a projection at every $k$.

Addressing PL maximization, the corresponding proximal mapping is derived in Appendix B

$$\text{prox}_{t_k g_{i_k}}(x_k) = x_k + t_k(N_0\exp(-c_k^*) - y_{i_k})a_{i_k} \quad \text{for } i_k = 1,\ldots, m,$$

where $c_k^*$ is found by implicit solution of

$$c_k^* = a_{i_k}^T x_k + t_k\|a_{i_k}\|_2^2(N_0\exp(-c_k^*) - y_{i_k}). \quad (20)$$

We also provide methods for calculating $c_k^*$ in Appendix B. For WLSQ, and LSQ, the proximal mapping can be derived by direct computation, shown in Appendix C,

$$\text{prox}_{t_k g_{i_k}}(x_k) = x_k - \frac{a_{i_k}^T x - b_{i_k}}{\|a_{i_k}\|^2 + (t_k w_{i_k})^{-1}} a_{i_k} \quad \text{for } i_k = 1,\ldots, m.$$

This WLSQ update step resembles the algebraic reconstruction technique (ART) with a relaxation parameter $\rho$. The difference is an additional term in the denominator which includes the weights $w_i$ and step-size $t_k$. In the discussion of the results, we label each of the instances of R-IPG1 by the corresponding optimization problem: TV-constrained (TVC) - PL maximization or TVC-PL, TVC-WLSQ, and TVC-LSQ.

For the present work, we refer to a subiteration as an update step based in a single row of $A$ and an iteration as a full cycle through all of the data followed by a single TV projection. Because we are interested in using few iterations, we choose $\rho = 1$ so that all $x_k$ for which $k\%(m+1) = 0$ are guaranteed to have a total-variation of $\gamma$ or less.

## 3. RESULTS

To demonstrate the incremental algorithms with the various data fidelity objective functions, we conduct image reconstruction studies with noisy sparse-view projection data. The digital phantom shown in Fig. 1, consisting of a 256×256 pixel array, models a slice of a breast for the dedicated breast CT application. Because we wish to isolate the issue of the impact of noise, we do not introduce additional data inconsistency, and the projection data are generated using the same linear operator as that used in the image reconstruction. The projection data consist of 100 projections onto a 512-bin linear detector

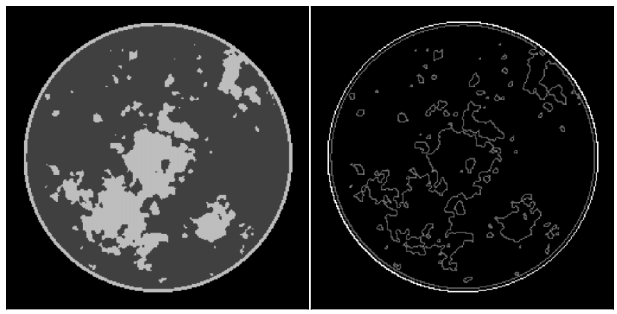

FIG. 1. Breast phantom for CT and its corresponding GMI. The fat and fibro-glandular tissue has linear attenuation coefficients of 0.194 $cm^{-1}$ and 0.233 $cm^{-1}$, respectively. Left is the linear attenuation map of the phantom in the gray scale window [0.174,0.253] $cm^{-1}$. Right is the GMI [0.0,0.1] $cm^{-1}$ that serves to illustrate that the test phantom is sparse in the GMI.

array. The source–detector and source-isocenter distances are modeled to be 72 and 36 cm, respectively. The sampling of this configuration is clearly not sufficient for direct or implicit inversion of Eq. (1) because the number of image pixels exceeds the number of measurements.

### 3.A. Noisy study

For all of the results, image reconstruction is performed on simulated data including noise. The same discrete-to-discrete model [Eq. (1)] generates the mean sinogram, and noise realizations for the transmitted x-ray intensity employ the Poisson model in Eq. (2). The integrated incident number of photons per view per detector bin is modeled to be uniform at two intensity levels: $N_0 = 2\times10^4$ and $2\times10^5$. The latter value corresponds to a fairly low intensity that might be used in an actual breast CT scan. In this simulation the only source of inconsistency is due to the noise model, and we know exactly what probability distribution function governs the noise realization selection. All other sources of inconsistency: continuous object model, beam hardening, scatter, partial volume averaging, etc., which would be present in actual CT data are suppressed. In this way, we can isolate and address the issue of the impact of maximum likelihood data fidelity terms for sparse-view CT using the incremental algorithms with truncated iteration.

### 3.B. Incremental algorithm parameter optimization

In the implementation of TVC-PL, TVC-WLSQ, and TVC-LSQ, we employed a decreasing step-size

$$t_k = \frac{t_0}{(\lfloor k/(m+1)\rfloor)^\alpha}, \tag{21}$$

where $\lfloor\cdot\rfloor$ is the floor operation, and the parameter $\alpha \in (0,1]$ is the decay exponential designed according to Eq. (15). The R-IPG1 instances presented in Sec. 2.C depend on three algorithm parameters: $t_0$, $\alpha$, and $\rho$. We consider running R-IPG1 with one hundred iterations or less under the restriction that the images respect the TV constraint $\mathrm{TV}(x_k) \le \gamma$ after every iteration. This restriction forces $\rho = 1$, and for such low iteration numbers the impact of $\alpha$ is weak. Accordingly, we set $\alpha = 0$, and consider only variation in $t_0$. Each of the optimization problems depends on the TV constraint parameter $\gamma$ and, in the case of TVC-WLSQ, on the choice of weights $w_i$. For the present study, we assume access to the phantom TV, $\gamma_0$, and for TVC-WLSQ, we assume knowledge of the object dependent variance in Eq. (10). While such knowledge is not realistic in actual application with unknown subjects, it serves to illustrate upper bound performance of the algorithms of interest.

Because the image TV constraint is satisfied at every iteration, it is sufficient to optimize $t_0$ based only on the corresponding value of the data fidelity objective function. Shown in Fig. 2 is the average data fidelity as a function of $t_0$ at iteration numbers 20, 50, and 100 for the three algorithms of interest setting $\gamma = \gamma_0$. The optimal value of $t_0$ is the one which minimizes the particular data fidelity for a given number of iterations. In all cases, optimal $t_0$ occurs at smaller values

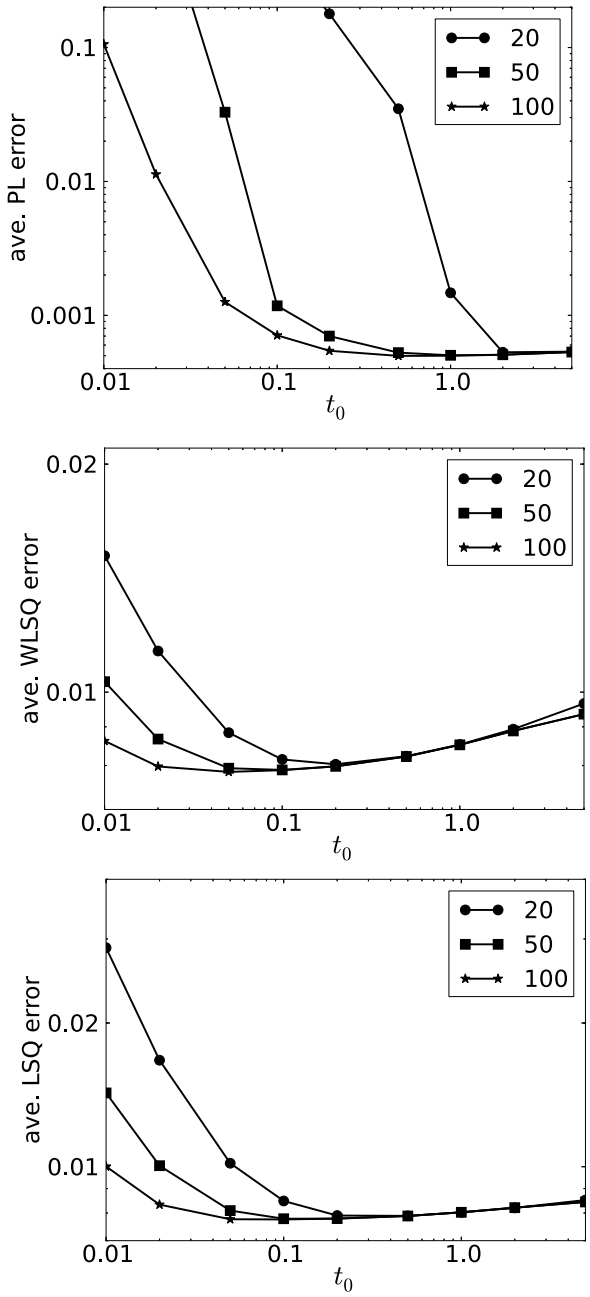


FIG. 2. Log–log plots of the average value of the data fidelity objective function for TVC-PL, TVC-WLSQ, and TVC-LSQ as a function of the incremental algorithm parameter $t_0$. Different curves in each graph correspond to different numbers of iterations, and these numbers appear in the legend. The average objective values are computed by evaluating the respective objective functions for the final image $x_k$ and dividing by the number of measurements $m$. In the case of the PL objective, the minimum possible value is not zero. As a result, the plotted average PL data-discrepancy represents the average over the PL data fidelity minus this fidelity evaluated with the phantom itself. It is also important to note that only the numbers within each graph can be compared; it is not meaningful to compare the values of different data fidelity objective functions.

when more iterations are performed. The curves for TVC-WLSQ and TVC-LSQ are fairly similar, while those of TVC-PL show rapid increase as $t_0$ decreases. The TVC-WLSQ and TVC-LSQ algorithms appear to be less sensitive to $t_0$ than TVC-PL.

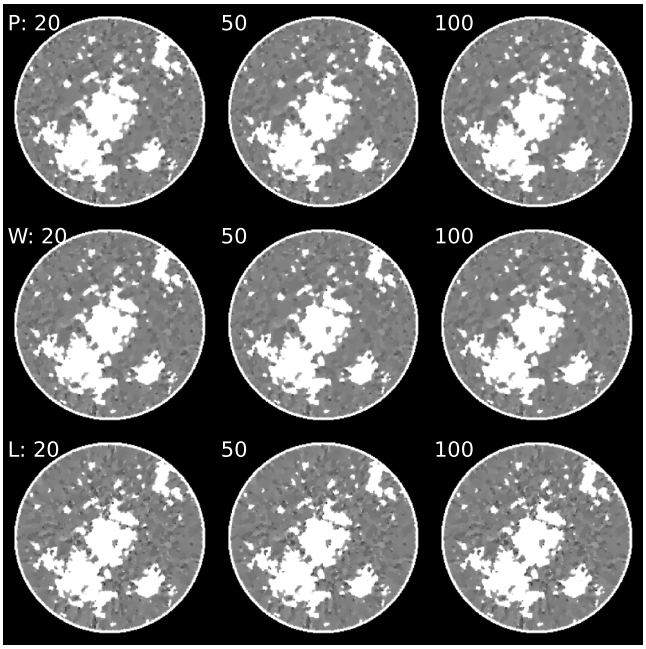


FIG. 3. Single images corresponding to the minima of each curve in Fig. 2. The top, middle, and bottom rows correspond to TVC-PL (denoted "P"), TVC-WLSQ (denoted "W"), and TVC-LSQ (denoted "L"), respectively. The left, middle, and right columns are the images at iteration number 20, 50, and 100, respectively. The display window is selected to be [0.184,0.204] $cm^{-1}$, centered on the background fat tissue. In this way, the noise level on the fat portion of the image is clearly depicted.

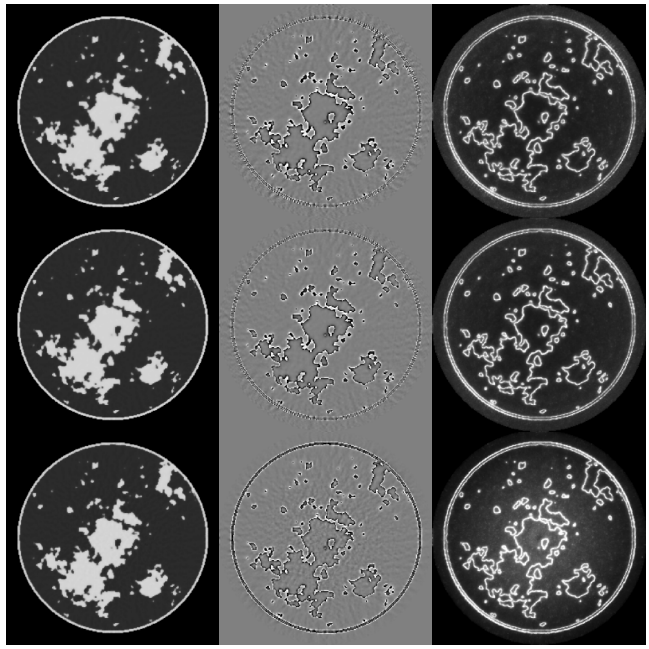

FIG. 4. Statistical properties of reconstructed images by TVC-PL (top row), TVC-WLSQ (middle row), and TVC-LSQ (bottom row) estimated from the images at 50 iterations for 1000 noise realizations of the data model. The left column is the mean image in the gray scale [0.184,0.243] $cm^{-1}$. The middle column is the breast phantom subtracted from the mean image in the gray scale [−0.005,0.005] $cm^{-1}$. Finally, the right column is the estimated image standard deviation also in the gray scale window [0.0,0.005] $cm^{-1}$.

For obtaining reconstructed images, we record the optimal value of $t_0$ for each algorithm at 20, 50, and 100 iterations. In Fig. 3, single images are shown for each case and for $N_0 = 2 \times 10^5$. Interestingly, even at 20 iterations the reconstructed images in all cases resemble the test phantom without any obvious artefacts due to truncated iteration. There is a slight sharpening of the TVC-PL images in going from 20 to 50 iterations; otherwise there is no clear difference between images for different iteration numbers. The images corresponding to TVC-PL and TVC-WLSQ appear to have less noise particularly toward the middle of the image. This is an indication that exploiting maximum likelihood does potentially have a beneficial impact in reducing noise variance without sacrificing bias.

### 3.C. Image statistics

To investigate the noise properties in more depth, we focus on the 50 iteration case and perform 1000 noise realizations using the Poisson model for the x-ray transmission data and reconstruct images with TVC-PL, TVC-WLSQ, and TVC-LSQ. From the 1000 reconstructed images a mean image is estimated along with a map of the pixel standard deviations all shown in Fig. 4 for the case of $N_0 = 2 \times 10^5$. The mean image is quite close to the test phantom, but subtracting the phantom from the mean images does highlight some differences between the results of the three algorithms. The standard deviation images are interesting in that the distributions are quite nonuniform as the standard deviation appears to be amplified near the borders between tissue types. A similar phenomenon was observed in Ref. 11 with the use of edge-preserving regularization. Also, the standard deviation of the image ensembles reconstructed by TVC-PL and TVC-WLSQ appears to have lower amplitude than that of TVC-LSQ particularly toward the middle of the image. This result strengthens the observation from the single images shown in Fig. 3 that use of maximum likelihood leads to lower amplitude noise in the reconstructed image than that corresponding to the unweighted data fidelity term.

To further investigate the reconstructed image noise properties, we generate standard deviation-bias curves by varying the TV constraint parameter $\gamma$ and computing the mean, standard deviation, and bias over the corresponding images. The results for both intensity settings are shown in Fig. 5. For $N_0 = 2 \times 10^4$, the TVC-WLSQ and TVC-PL curves are similar with both results having lower variance and bias than the corresponding points of the TVC-LSQ results. The results for $N_0 = 2 \times 10^5$ appear quite different. The TVC-WLSQ and TVC-LSQ curves show a minimum average bias at $\gamma = \gamma_0$. While for TVC-PL the minimum average bias is obtained when $\gamma = 0.95\gamma_0$. The average image standard deviations are quite close for TVC-PL and TVC-WLSQ, indicating that the Gaussian approximation of the Poisson model and the linearization of the data variance through the negative natural logarithm operation are quite good approximations. The TVC-LSQ curve shows significantly higher standard deviation than the other two for all selected $\gamma$. There is a surprisingly large

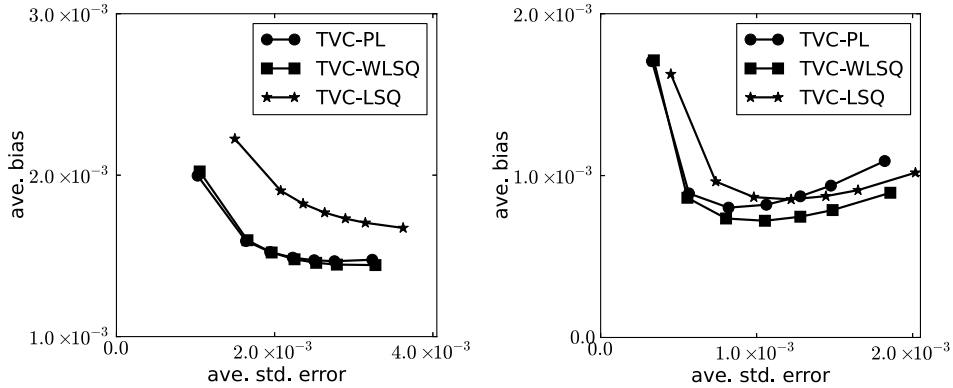


FIG. 5. Average bias versus average standard deviation curves generated by image reconstruction with each of the incremental algorithms at 50 iterations for (left) $N_0 = 2\times10^4$ and (right) $N_0 = 2\times10^5$. The curves are generated by varying the TV constraint parameter $\gamma$ among seven values. These TV values are computed as the fractions $[0.8, 0.9, 0.95, 1.0, 1.05, 1.1, 1.2]$ of the phantom TV, $\gamma_0$. The curve values are indicated by the various symbol dots with the one corresponding to the lowest fraction, 0.8, appearing on the leftmost point of each curve.

gap between the average bias of TVC-PL and TVC-WLSQ with the latter below the former.

In order to determine whether or not the gap between TVC-PL and TVC-WLSQ for the $N_0 = 2\times10^5$ case is due to a discrepancy in convergence, we estimate image statistics from an ensemble of 100 images, which are converged solutions to the respective optimization problems. The converged images are obtained from 2000 iterations of the CP (Refs. 9 and 10) algorithm, and the resulting plots are shown in Fig. 6. For the case of $N_0 = 2\times10^4$, the agreement between the R-IPG1 and CP results is close and both sets of results show similar curves. For $N_0 = 2\times10^5$, however, there is a larger discrepancy between the R-IPG1 and CP results. The CP results show close agreement between CP-TVC-WLSQ and CP-TVC-PL, while the R-IPG1 results do not. Thus, the latter disagreement can be attributed to slower convergence of TVC-PL relative to TVC-WLSQ.

Another interesting phenomenon in the $N_0 = 2\times10^5$ results of Fig. 5 is that both the standard deviation and bias increase for each curve as $\gamma$ increases beyond $\gamma_0$. In order to see why this happens, we focus on mean and standard deviation maps, shown in Fig. 7, for TVC-WLSQ for three different values of $\gamma$. For the case of $\gamma = 0.8\gamma_0$ the standard deviation map has clearly lower amplitude than that of $\gamma = \gamma_0$, but the bias is visually larger as the mean image is significantly blurred for the lower value of $\gamma$. For the $\gamma = 1.2\gamma_0$ results, the standard deviation map has larger amplitude as expected. The bias map in this case reflects a situation unique to view angle under-sampling. For such conditions, the TV constraint not only serves to regularize the image against noise, but it also assists the image recovery in the noiseless case. As the TV constraint is loosened the phantom is no longer recovered from ideal under-sampled data. As a result streak artefacts become apparent in the bias map and the average bias rises relative to that of $\gamma = \gamma_0$. This effect is less pronounced in the $N_0 = 2\times10^4$ results because data noise is more dominant.

## 4. PRACTICAL ADVANTAGES OF EACH OF THE R-IPG1 ALGORITHMS

First, we point out that the performance of TVC-PL, TVC-WLSQ, and TVC-LSQ is remarkable considering the simple structure of these algorithms. Over the 1000 noise realizations each algorithm achieved the TV constraint to within 0.01% of the target $\gamma$ and the final data fidelity varied less than 0.1% over the ensemble of realizations. Most importantly, the

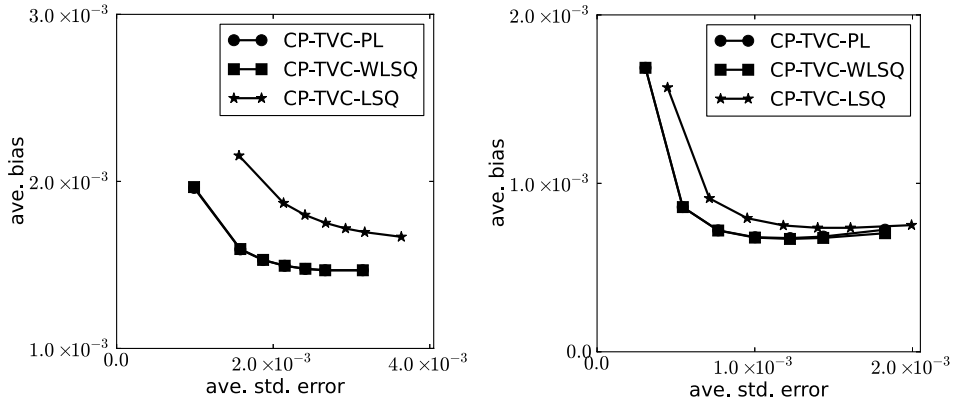


FIG. 6. Same as Fig. 5 except that curves are generated from an ensemble of 100 images obtained by solving the respective optimization problems with the CP algorithm.

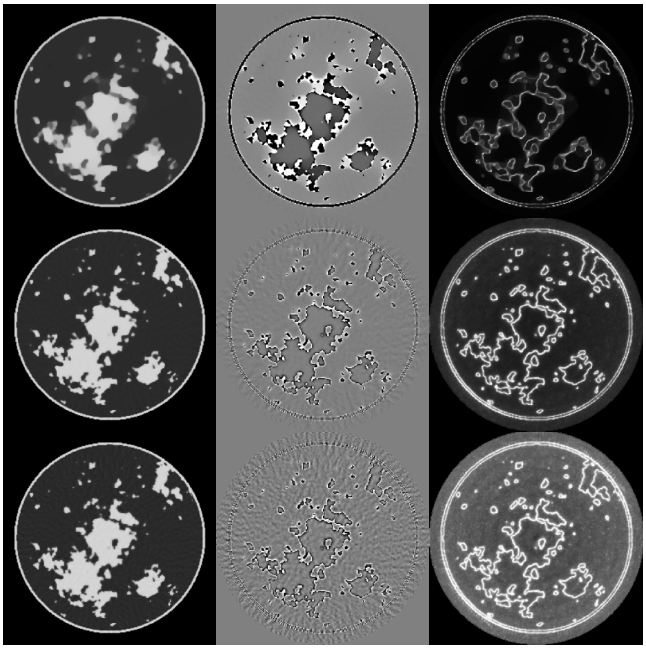

FIG. 7. Same as Fig. 4 except that the rows all correspond to 50 iterations of TVC-WLSQ for the TV constraint parameters: $\gamma = 0.8\gamma_0$ (top row), $\gamma = \gamma_0$ (middle row), and $\gamma = 1.2\gamma_0$.

visual appearance of the reconstructed images is close to the test phantom for low numbers of iterations.

Use of TVC-PL appears to reduce noise variance in the reconstructed image relative to TVC-LSQ provided that the noise model matches the Poisson distribution of the x-ray transmission data. TVC-PL is slightly more complicated to implement than TVC-LSQ, but the difference in implementation and algorithm efficiency is minor. Moreover, TVC-PL does not introduce more optimization parameters than TVC-LSQ. As a result, if the assumed data model applies, or is a close approximation, TVC-PL provides the better option to obtain images with reduced noise amplitude without sacrificing image bias.

In the performed simulation studies, it appears at first glance that TVC-WLSQ is the best option even though the data realizations are generated with the Poisson model. We point out, however, that the achieved performance in terms of bias versus standard deviation assumed knowledge of the ideal weights for the quadratic data fidelity. In practice, such weights are not available. If the existing data are used to obtain an estimate for the weights, the advantage in terms of image variance will be lost because the existing data contain noise. A smoothed version of the data may provide a better estimate of the weights. Also, the TVC-WLSQ algorithm may have further advantage when other sources of noise are considered such as detector electronic noise.

The TVC-LSQ algorithm may also have advantages in some situations. As pointed out noise is not the only form of data inconsistency, and if other physical factors are more important than the difference in noise modeling then TVC-LSQ is a good option. It appears also that the convergence of TVC-LSQ is slightly faster than that of TVC-PL at least for the CT system investigated here.

## 5. CONCLUSION

We have applied the incremental framework of Refs. 6 and 7 to generate algorithm instances of TV-constrained, data-discrepancy minimization for CT image reconstruction. The algorithm yields an accurate solution to the designed optimization problem at large iteration numbers and can provide a useful image at low iteration numbers. The TVC-PL, TVC-WLSQ, and TVC-LSQ are all run at low iteration numbers, yet the resulting images appear to reflect the design considerations of the corresponding optimization problem. Each of the optimization problems contains the TV constraint for mitigating streaking due to angular under-sampling. Indeed, each of the reconstructed images appear to be free from this artefact with $\gamma \le \gamma_0$. Furthermore, the results of the noise studies with each algorithm indicate that the maximum likelihood data fidelity terms have an impact even though the iteration number is severely truncated.

## ACKNOWLEDGMENTS

M.S.A. was supported by Grant No. 291405 "HD-Tomo" from the European Research Council. This work was also supported in part by NIH R01 Grant Nos. CA158446, CA120540, EB000225, CA182264, and EB018102. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

## APPENDIX A: PROXIMAL MAPPING OF DISTANCE FROM A CONVEX SET

Let $g(x) = \lambda d_C(x)$. Our goal is to find

$$\text{prox}_{tg}(x) = \arg\min_u \{\lambda d_c(u) + \frac{1}{2t}\|u-x\|_2^2\}. \tag{A1}$$

We shall use the optimization condition for unconstrained convex minimization that $u^*$ is an optimal point if and only if $0 \in \partial_c(u^*)$. For proof of this fact we refer the reader to Ref. 12. We thus need the subdifferential of $d_c(x)$ which is given by

$$\partial d_c(u) = \begin{cases} N_C(u) \cap \{v : \|v\|_2 \le 1\} & \text{if } u \in C \\ \left\{ \dfrac{u - \mathcal{P}_C(u)}{\|u - \mathcal{P}_C(u)\|_2} \right\} & \text{else} \end{cases}, \tag{A2}$$

where for a convex set, the normal cone, $N_C(u)$ can be written as

$$N_C(u) = \{y : y^T(z-u) \le 0 \ \ \forall z \in C\}, \tag{A3}$$

we thus see that at an optimal point $u^*$, one of two conditions must hold

(1) $x - u^* = \dfrac{\lambda t}{d_c(u^*)}(u^* - \mathcal{P}_c(u^*)), \ \ u^* \notin C$

(2) $x - u^* \in \lambda t N_C(u^*) \cap \{v : \|v\|_2 \le 1\}, \ \ u^* \in C.$

If condition (1) holds, we see that $x$, $u^*$, and $\mathcal{P}_C(u^*)$ all fall along a line, which, along with $u^* \notin C$ and the convexity of

$C$, implies that $\mathcal{P}_C(u^*) = \mathcal{P}_C(x)$. In addition, if $\lambda t < d_C(x)$, we see that $u^*$ is a convex combination of $x$ and $\mathcal{P}_C(x)$, namely,

$$u^* = x + \lambda t \frac{\mathcal{P}_C(x) - x}{d_C(x)} \quad \text{if } d_C(x) > \lambda t. \tag{A4}$$

On the other hand, if condition (2) holds, then we have

$$\|x - u^*\|_2 \le \lambda t, \tag{A5}$$

$$(x - u^*)^T (z - u^*) \le 0 \quad \forall \quad z \in C. \tag{A6}$$

It is clear that $\mathcal{P}_C(x)$ satisfies both conditions assuming $\lambda t \ge d_C(x)$. We have thus found points satisfying the optimality condition in two cases: when $\lambda t < d_c(x)$ and when $\lambda t \ge d_c(x)$. Since the minimization is performed on a strongly convex function, we know these points are the unique minima of the function in the two cases. We have thus obtained the desired result.

## APPENDIX B: PROXIMAL MAPPING FOR LINEAR PLUS EXPONENTIAL

Let $g(x) = ya^T x + N\exp(-a^T x)$, where $y \in \mathfrak{R}$, $a \in \mathfrak{R}^n$, $x \in \mathfrak{R}^n$, and $N \in \mathfrak{R}$. Our goal is to find

$$\text{prox}_{tg}(x) = \arg\min_u \{g(u) + \frac{1}{2t}\|u - x\|_2^2\}. \tag{B1}$$

Noting that the argument is everywhere differentiable, we may simply set the gradient equal to zero to obtain our optimality condition, yielding

$$ya - N\exp(-a^T u^*)a + t^{-1}(u^* - x) = 0. \tag{B2}$$

Dotting both sides with $a$ and rearranging terms we find

$$a^T(u^* - x) = t\|a\|_2^2(N\exp(-a^T u^*) - y). \tag{B3}$$

We suggest two methods to solve for $a^T u^*$ in Eq. (B3). First, one can use a root finding method such as the Newton–Raphson method. Second, one can rearrange Eq. (B3) further to yield

$$(a^T(u^* - x) + t\|a\|_2^2 y)e^{a^T(u^* - x) + t\|a\|_2^2 y} = t\|a\|_2^2 N e^{t\|a\|_2^2 y - a^T x}, \tag{B4}$$

from which one can see the solution is

$$a^T u^* = W_0(t\|a\|_2^2 N e^{t\|a\|_2^2 y - a^T x}) - t\|a\|_2^2 y + a^T x, \tag{B5}$$

where $W_0$ is the Lambert $W$-function (we confine our attention here to the real-valued principle branch $W_0$) which is defined to be the inverse of the function,

$$g(z) = ze^z, \tag{B6}$$

and is a single-valued mapping of the non-negative reals. This solution may be useful because methods for calculating the Lambert $W$-function are provided in many standard programming libraries.

Denoting the solution to Eq. (B3) as $c^* = a^T u^*$, we then have from Eq. (B2),

$$\text{prox}_{tg}(x) = x + t(N\exp(-c^*) - y)a. \tag{B7}$$

## APPENDIX C: PROXIMAL MAPPING OF WEIGHTED QUADRATIC

Let $g(x) = (a^T x - b)^2 w$, where $a \in \mathfrak{R}^n$, $b \in \mathfrak{R}$, and $w \in \mathfrak{R}$. Our goal is to find

$$\text{prox}_{tg}(x) = \arg\min_u \{(a^T u - b)^2 w + \frac{1}{2t}\|u - x\|_2^2\}. \tag{C1}$$

Since the argument is differentiable everywhere, our optimality condition is found simply by setting the gradient to zero, yielding

$$w(a^T u^* - b)a + \frac{1}{t}(u^* - x) = 0. \tag{C2}$$

Dotting both sides with $a$ and rearranging terms, we find

$$a^T u^* = \frac{a^T x + b\|a\|_2^2 wt}{1 + \|a\|_2^2 wt}. \tag{C3}$$

Plugging this result back into the original expression yields

$$u^* = x - \frac{a^T x - b}{(wt)^{-1} + \|a\|_2^2}, \tag{C4}$$

which is the desired result.

[a]Electronic mail: seanrose949@gmail.com
[b]Electronic mail: mskan@dtu.dk
[c]Electronic mail: sidky@uchicago.edu
[d]Electronic mail: xpan@uchicago.edu

[1]R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," J. Theor. Biol. **29**, 471–481 (1970).
[2]J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," Phys. Med. Biol. **51**(15), R541–R578 (2006).
[3]E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).
[4]E. Y. Sidky, X. Pan, I. Reiser, R. M. Nishikawa, R. H. Moore, and D. B. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," Med. Phys. **36**, 4920–4932 (2009).
[5]E. Y. Sidky, Y. Duchin, X. Pan, and C. Ullberg, "A constrained, total-variation minimization algorithm for low-intensity X-ray CT," Med. Phys. **38**, S117–S125 (2011).
[6]D. P. Bertsekas, "Incremental proximal methods for large scale convex optimization," Math. Program. **129**, 163–195 (2011).
[7]M. S. Andersen and P. C. Hansen, "Generalized row-action methods for tomographic imaging," Numer. Algorithms **67**, 121–144 (2014).
[8]S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero, "Convergent incremental optimization transfer algorithms: Application to tomography," IEEE Trans. Med. Imaging **25**, 283–295 (2006).
[9]A. Chambolle and T. Pock, "A first-order primal–dual algorithm for convex problems with applications to imaging," J. Math. Imaging Vis. **40**, 120–145 (2011).
[10]E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," Phys. Med. Biol. **57**, 3065–3091 (2012).
[11]T. Köhler and R. Proksa, "Noise properties of maximum likelihood reconstruction with edge-preserving regularization in transmission tomography," in *Proceedings of the 9th International Meeting on Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* (Beijing, China, 2009), pp. 263–266, available at http://www.fully3d2009.org.
[12]D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization* (Athena Scientific, Nashua, NH, 2003).